Deep reinforcement learning and simulation as a path toward precision medicine

CASIS 2018

Brenden Petersen, Jiachen Yang, Will Grathwohl,
Chase Cockrell, Claudio Santiago, Thomas Desautels,
Gary An, Dan Faissol

LLNL-PRES-751582
This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC

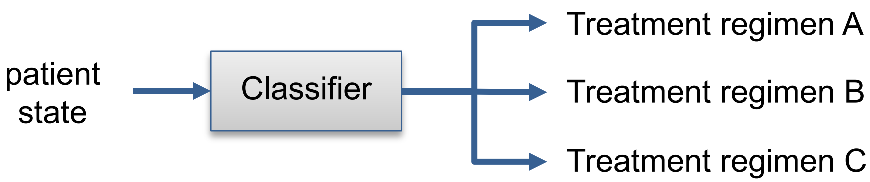
Lawrence Livermore National Laboratory

Precision medicine as a control problem

Traditional precision medicine

Classify then *treat*

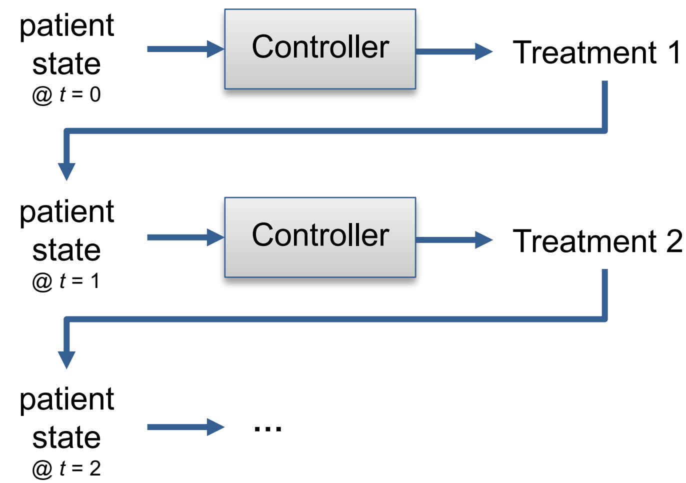
> "…the ability to classify individuals into subpopulations that differ in their susceptibility to a particular disease or their response to a specific treatment."
> - National Research Council



- Viewed as a classification task
- Therapies are <u>static</u> and <u>non-adaptive</u>

Proposed vision

Dynamic, feedback control



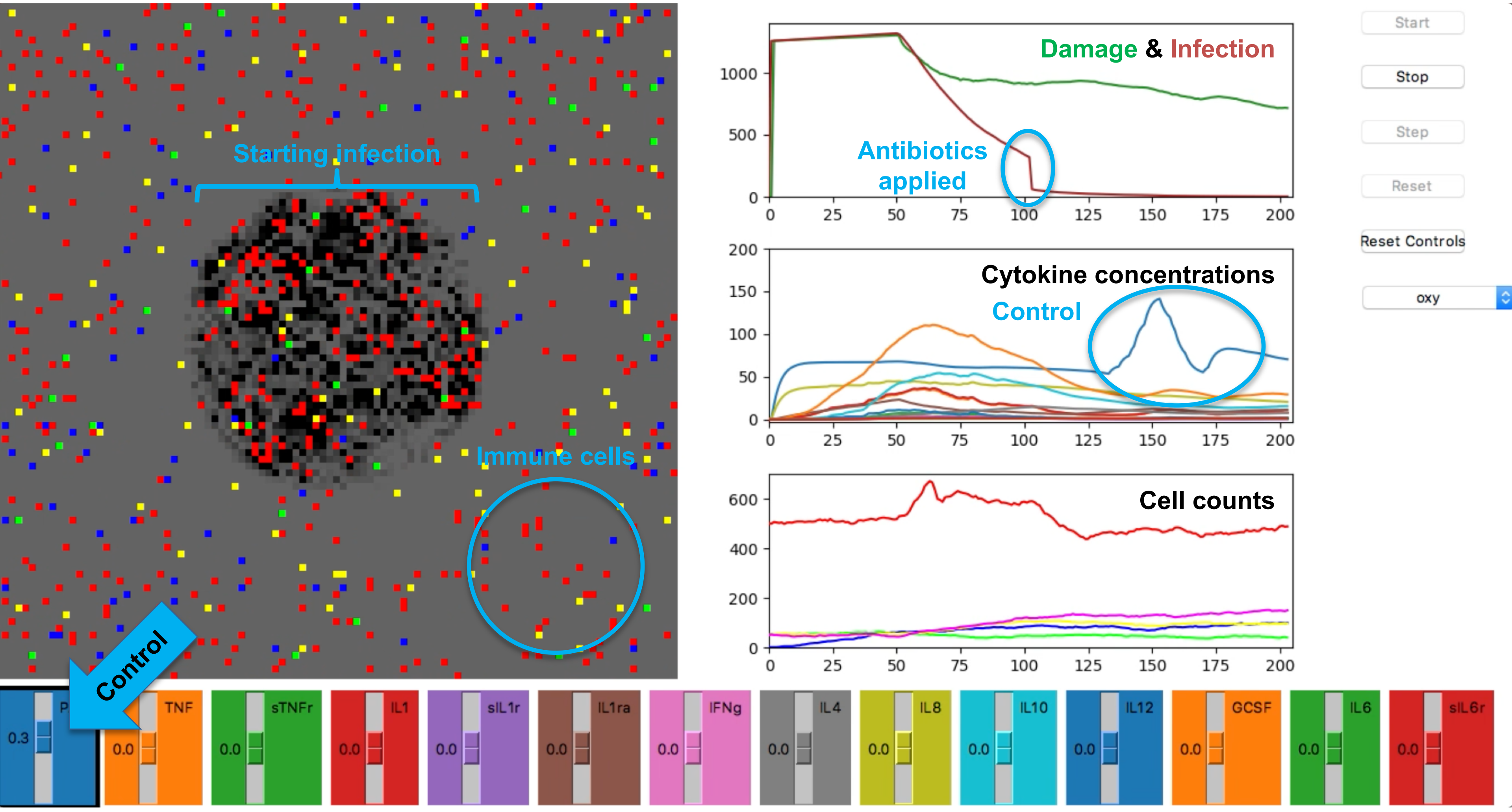
- Viewed as an optimal control task
- Therapies are <u>dynamic</u> and <u>adaptive</u>
  - Dependent upon patient trajectory

The need for simulation

- Many control approaches use existing data to retrospectively learn control policies
- Simulation enables virtual experimentation: going beyond what has been tried
- Recent advances in optimal control have enabled learning controllers for complex, high-dimensional simulations

	Learning controllers using…	
	Clinical Data	**Biological Simulation**
Scope of interventions	Limited to what's already been tried	Able to explore new interventions and/or combinations
Interpretability of interventions	Limited by statistical power of existing data	Limited only by computation
Dimensionality of interventions	Low-dimensional, discrete (e.g. 1 – 2 drugs, 3 doses)	High-dimensional, continuous
Dynamics of interventions	Typically static	Dynamic, adaptive

Sepsis agent-based simulation – Demo



Reinforcement learning (RL)

Given an

observation

choose the

action

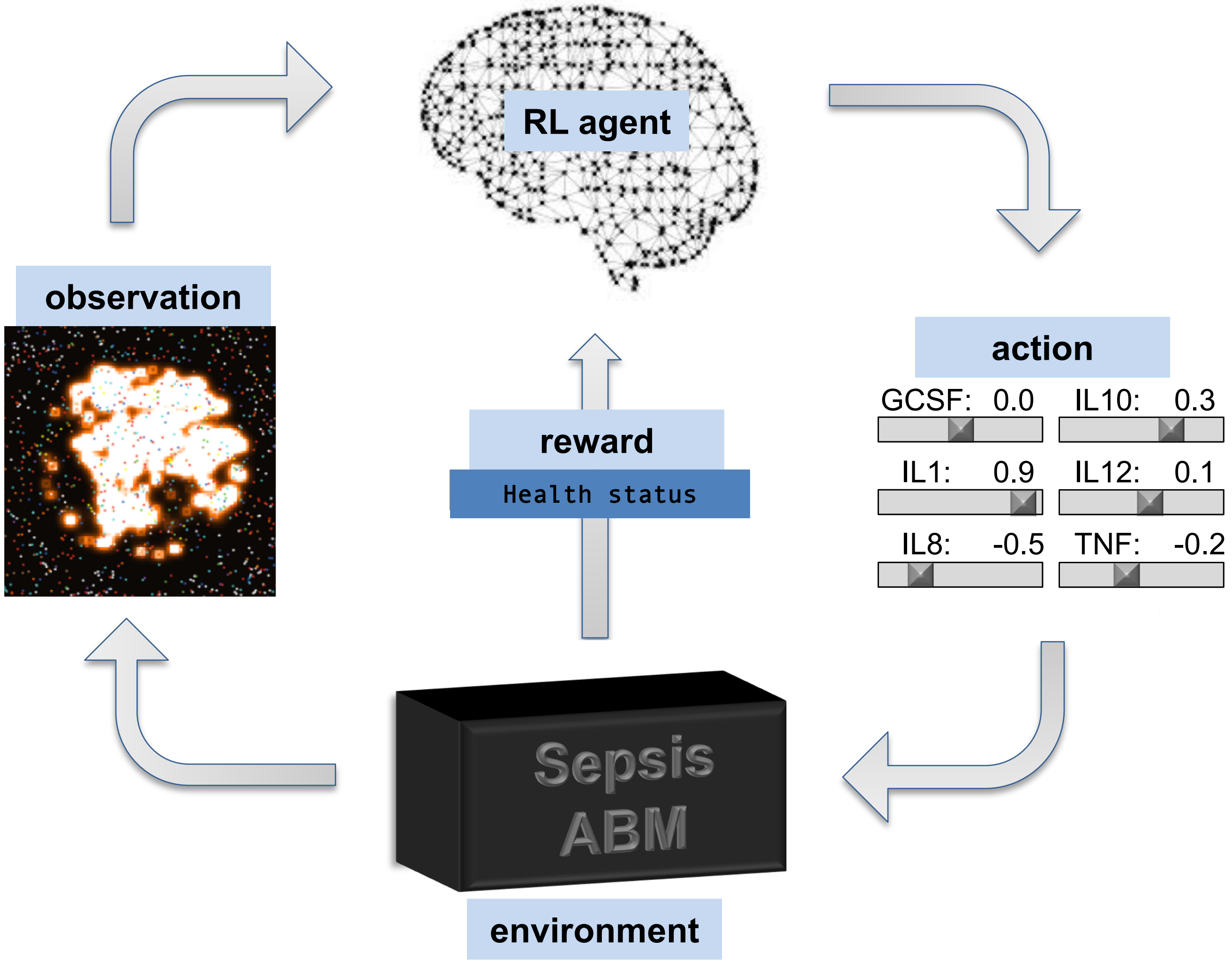
expected to maximize the cumulative

reward

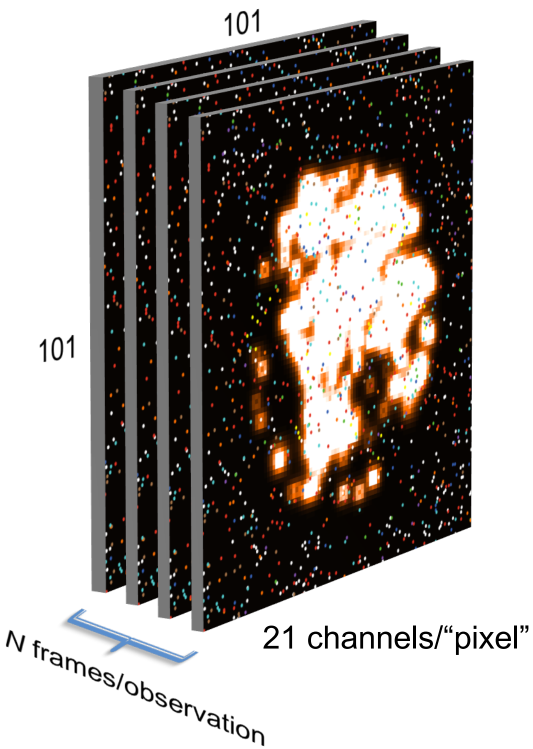
RL agent

learns by interacting with the

environment



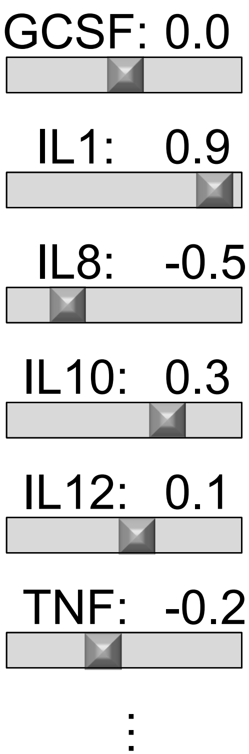
Problem Formulation: Observation Space



Observation Space

large, spatial	small, aggregate
Cytokine level + cell counts at each grid point	Aggregate cytokine levels + cell counts (non-spatial)
Size: $\mathbb{R}^{101 \times 101 \times 21 \times N}$	Size: $\mathbb{R}^{21 \times N}$
Clinically unrealistic with today’s technology	Clinically plausible from blood tests

Problem Formulation: Action Space



Action Space

large, continuous	**small, discrete**
Differentially control all cytokines at once	Augment or inhibit by a fixed amount; One cytokine at a time
Size: $[-1, 1]^{14}$	Size: 29
Clinically plausible with multi-channel infusion pump	Clinically plausible

observation
action
reward

Problem Formulation: Reward Signal

- The simulation naturally provides only sparse, binary rewards: life/death

$$r_{\text{outcome}} = \lambda_{+}[\text{heal}] - \lambda_{-}[\text{die}]$$

- To aid learning, we added two terms to the reward signal
  1. Potential-based reward shaping term
     - Helps guide the RL agent toward "good" states without altering the optimal policy

$$r_{\phi} = \lambda_{\phi}\left(\text{damage}(s) - \text{damage}(s')\right)$$

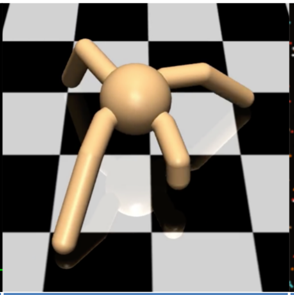
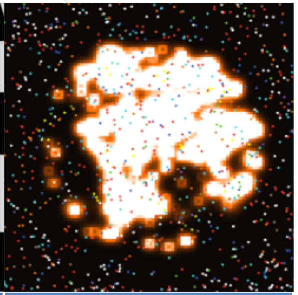
  2. A penalty for taking actions
     - Regularizer; promotes conservative actions

$$r_a = -\lambda_a \|a\|_1$$

- Final reward signal: $r(s, a, s') = r_{\text{outcome}} + r_{\phi} + r_a$

Unique challenges of the sepsis environment

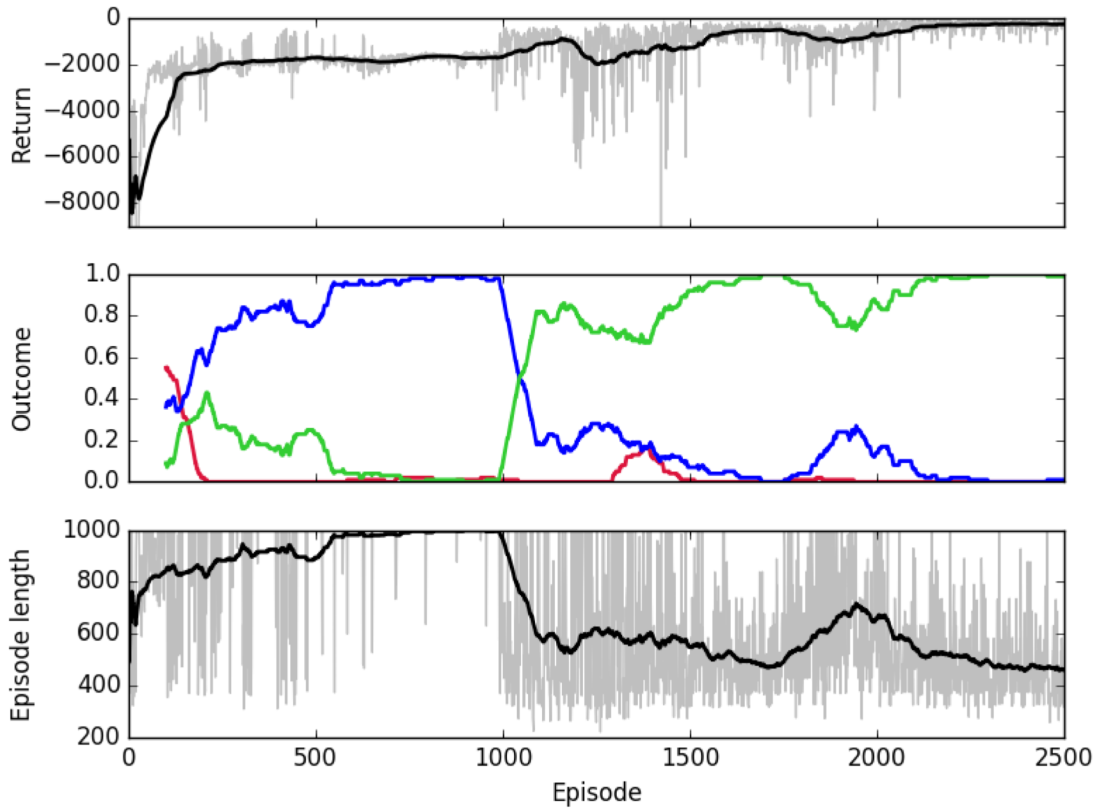
Failed to solve using human experience, genetic algorithms, and classify → control approaches

Challenge	Atari 2600	MuJoCo	Sepsis
High-dimensional state	✓	✓	✓
High-dimensional actions		✓	✓
Sparse rewards	sometimes		✓
Long time horizons			✓
Computationally expensive			✓
Unsolvable by humans			✓
Stochastic	None	None	**High**
Each episode has different dynamics			✓

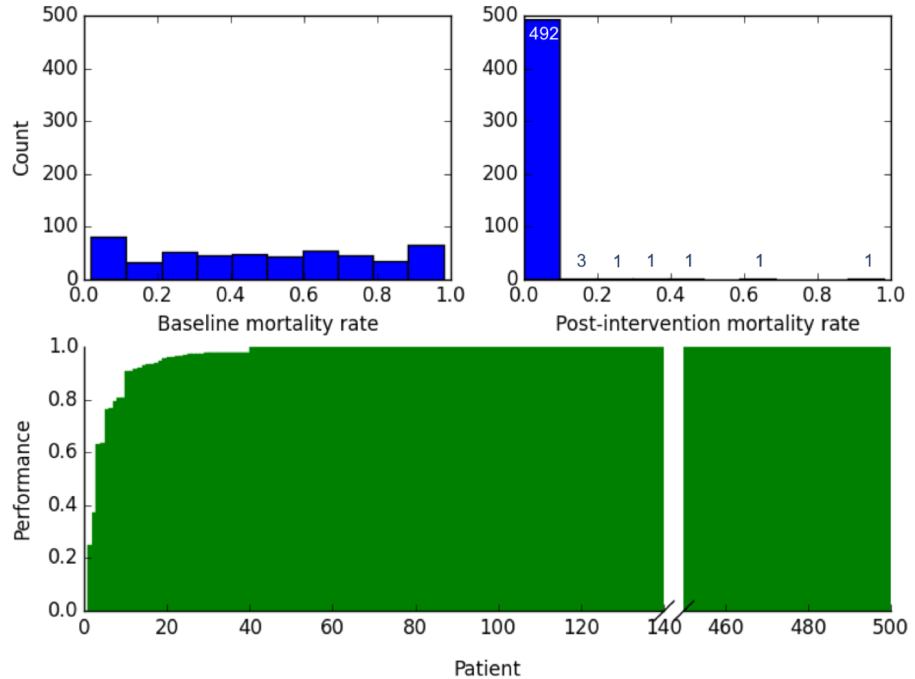
Training the DRL agent

- Environment is "solved" by 2,500 episodes
- Distinct "phases" of learning

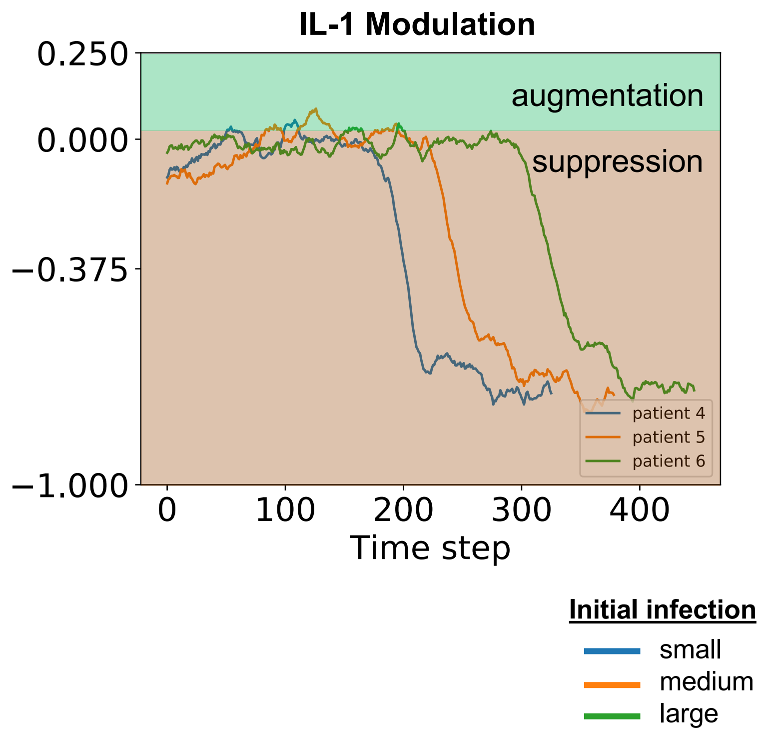


Evaluating the learned policy

- Mortality rate under learned policy
  - Trained patient: 46% → 0%
  - Across 500 patients: 49% → 0.8%



- Clinical insight
  - IL-1 (pro-inflammatory) is unregulated early and suppressed late
  - Suppression comes later for patients with larger initial infections

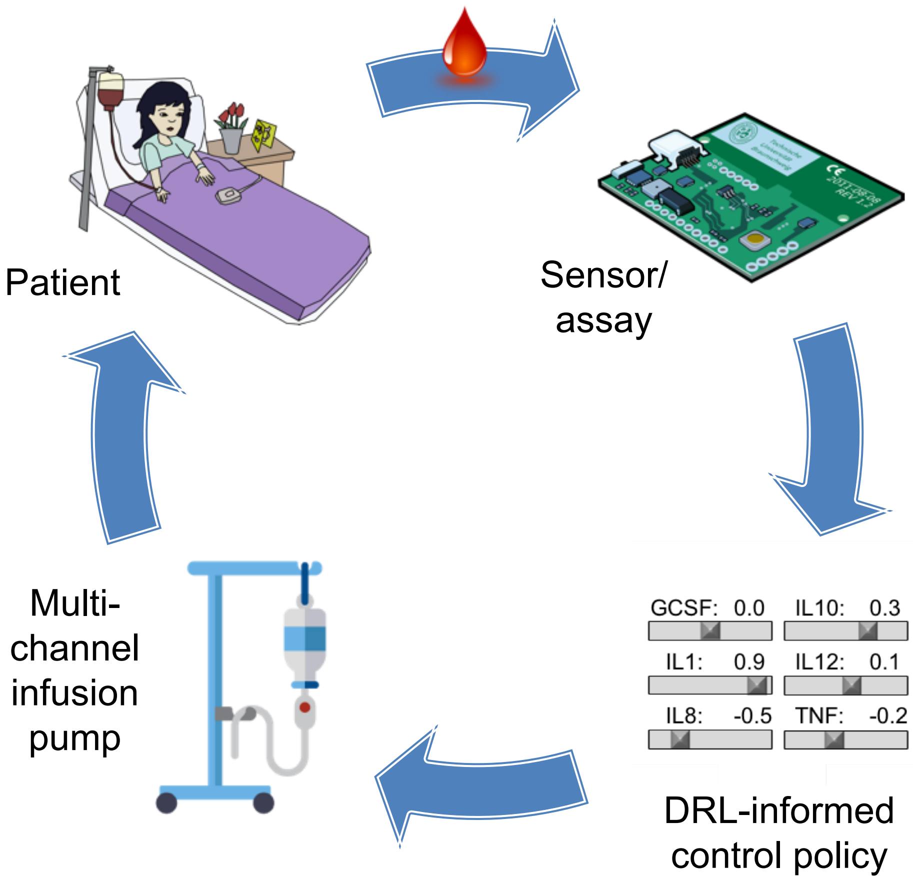


Next steps: Improving clinical plausibility

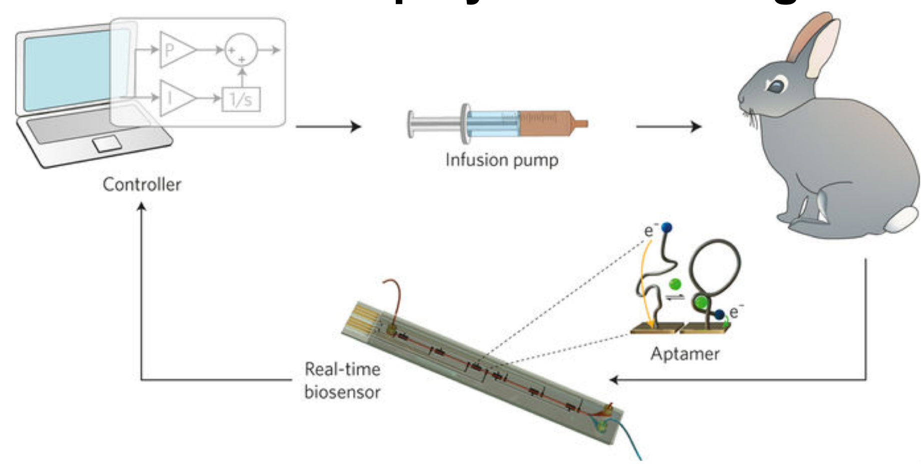
- Tradeoff between *controllability* and *clinical relevance*

Aspect	*Clinically implausible*		*Clinically plausible*	**Next step**
Observability	spatially resolved; all state variables		concentrations; cytokines, cells	Remove infection and damage from state
Observation delay	theoretical: 0 hr	wet-lab: 3 hr	laboratory test: 24 – 48 hr	Add 3 hr observation delay
Observation frequency	invasive: ~minutes		clinical practice: 6 hr	Decrease to observation every 6 hr
Action set	theoretical drugs	wet-lab tested drugs	clinically tested drugs	Identify existing mediators

Long-term vision: Closed-loop control system



Closed-loop dynamic dosing



"…used to modulate the drug infusion rate to **maintain in vivo drug levels** at a target value."
- Nat Biomed Eng 1:72 (2017)

https://openclipart.org/
https://www.mediware.com/home-care/blog/new-legislation-help-home-infusion-patients/

Thank you!

See Tom Desautel's poster!